JOURNAL OF MICROBIOLOGY & BIOLOGY EDUCATION
DOI: https://doi.org/10.1128/jmbe.v18i3.1320

Research

Development, Validation, and Application of the Microbiology Concept Inventory †

Timothy D. Paustian[1]*, Amy G. Briggs[2], Robert E. Brennan[3], Nancy Boury[4], John Buchner[5], Shannon Harris[1], Rachel E. A. Horak[6], Lee E. Hughes[7], D. Sue Katz-Amburn[8], Maria J. Massimelli[9], Ann H. McDonald[10], Todd P. Primm[11], Ann C. Smith[5], Ann M. Stevens[12], and Sunny B. Yung[11]
[1]University of Wisconsin-Madison, Madison, WI 53706; [2]Beloit College, Beloit, WI 53511; [3]University of Central Oklahoma, Edmond, OK 73034; [4]Iowa State University, Ames, IA 50011; [5]University of Maryland, College Park, MD 20742; [6]American Society of Microbiology, Washington, DC 20036; [7]University of North Texas, Denton, TX 76203; [8]Rogers State University, Claremore, OK 74017; [9]University of California – Irvine, Irvine, CA 92697; [10]Concordia University Wisconsin, Mequon, WI 53097; [11]Sam Houston State University, Huntsville, TX 77340; [12]Virginia Tech, Blacksburg, VA, 24061

If we are to teach effectively, tools are needed to measure student learning. A widely used method for quickly measuring student understanding of core concepts in a discipline is the concept inventory (CI). Using the American Society for Microbiology Curriculum Guidelines (ASMCG) for microbiology, faculty from 11 academic institutions created and validated a new microbiology concept inventory (MCI). The MCI was developed in three phases. In phase one, learning outcomes and fundamental statements from the ASMCG were used to create T/F questions coupled with open responses. In phase two, the 743 responses to MCI 1.0 were examined to find the most common misconceptions, which were used to create distractors for multiple-choice questions. MCI 2.0 was then administered to 1,043 students. The responses of these students were used to create MCI 3.0, a 23-question CI that measures students' understanding of all 27 fundamental statements. MCI 3.0 was found to be reliable, with a Cronbach's alpha score of 0.705 and Ferguson's delta of 0.97. Test item analysis demonstrated good validity and discriminatory power as judged by item difficulty, item discrimination, and point-biserial correlation coefficient. Comparison of pre- and posttest scores showed that microbiology students at 10 institutions showed an increase in understanding of concepts after instruction, except for questions probing metabolism (average normalized learning gain was 0.15). The MCI will enable quantitative analysis of student learning gains in understanding microbiology, help to identify misconceptions, and point toward areas where efforts should be made to develop teaching approaches to overcome them.

INTRODUCTION

In 2012, the President's Council of Advisors on Science and Technology set a challenge to improve STEM education (1). Numerous high-impact practices are being employed to reach this goal, such as writing across the curriculum, cooperative learning, problem-based learning, and flipped classrooms, among many others (2–5). To support a concerted effort to determine what students are learning about microbiology, we worked to develop an assessment tool that would reveal student understanding of significant microbiology concepts. In 1992, Hestenes, Wells, and Swackhamer were struggling with a similar dilemma and developed the Force Concept Inventory (FCI) to assess students' understanding of concepts related to force in physics (6). This concept inventory was a short test consisting of multiple choice questions (MCQs) directed specifically at commonly held misconceptions about force. Use of the FCI as a pre- and post-course assessment tool allowed faculty to determine whether changes in student knowledge were moving toward a scientific understanding of force. The ability of the tool to assess student learning about concepts important to physicists caused the FCI to be widely employed and helped to catalyze a dramatic transformation in how physics is taught (7).

Following the success of the FCI, numerous groups have developed and employed Concept Inventories (CI). Each CI

*Corresponding author: Mailing address: Microbial Sciences Building, Room 2517, University of Wisconsin-Madison, 1550 Linden Drive, Madison, WI 53706. Phone: 608-263-4921. E-mail: paustian@wisc.edu.
Received: 13 March 2017, Accepted: 28 July 2017, Published: 5 October 2017.
†Supplemental materials available at http://asmscience.org/jmbe

addresses student misconceptions about a defined set of concepts, in a multiple-choice test that has been validated through trial runs and evaluation. Concept inventories are now available for astronomy and space science (8), relativity (9), digital logic (10), statistics (11), calculus (12), discrete mathematics (13), chemistry (14), biomechanics (15), general biology (16), central dogma (17), circulation (18), biology experimental design (19), flowering plant growth (20), diffusion and osmosis (21), genetics (22), meiosis (23), molecular biology (24), the lactose operon (25), and host-pathogen interactions (26).

Currently, there is not a general microbiology CI. The first step in developing such a tool would be to articulate the concepts that educators in the field found most important for student learning. In 2012, the American Society of Microbiology Task Force for Curriculum Guidelines developed a set of 27 fundamental statements, organized into six areas, concisely describing the core concepts that a well-educated microbiology student should understand (27). In 2014, learning outcomes, developed from these fundamental statements, were created to provide examples of activities microbiology students should be able to perform (28).

These widely vetted lists of important concepts for microbiology served as the foundation for the development of the Microbiology Concept Inventory (MCI), described here, and the Microbiology for Health Science Concept Inventory, described in the accompanying paper in this issue (29). In this paper, we report the development of the MCI and subsequent analysis to demonstrate that the instrument is valid and reliable in measuring students' conceptual understanding in microbiology. An accompanying paper in this issue (30) discusses common misconceptions students bring into microbiology courses, with suggestions on how to address them.

METHODS

The protocol for the development of the MCI was informed by the development of the Host-pathogen CI (26) and the protocol for test development outlined by Adams and Wieman (31). The steps included selection of concepts considered important by microbiology educators, identification of student thinking about the concepts and commonly held misconceptions, development of multiple-choice questions (MCQs) using these misconceptions as distractors, validation interviews on both novices and subject experts that were not involved with the instrument development, and finally, statistical validation of the concept inventory.

IRB

This study was completed in compliance with human subject IRB buchanan@beloit.edu (Beloit College); IRB 14194 (University of Central Oklahoma); IRB 16-166 (Iowa State University); IRB 750585-2 (University of Maryland); IRB 14-515 (University of North Texas); IRB 6663 (Rogers State University/University of Oklahoma); IRB 2016-2750 (University of California – Irvine); IRB 704284-2 (Concordia University Wisconsin); IRB 19512 (Sam Houston State University); IRB 11-017 (Virginia Tech); IRB 2014-1466, 2015-1272 (University of Wisconsin – Madison).

Data collection and confidentiality

In all classes, the study was described to students and informed consent obtained. Students were given the choice to participate or not, with the instructors not knowing the results of this decision. Collected data from all courses was anonymized by assigning random numbers to each response set. All data were combined before being distributed for analysis.

Selection of concepts for the MCI

Our efforts were accelerated by being able to use the ASMCG for Undergraduate Microbiology (32). Using the fundamental statements and learning outcomes from the ASM, the authors of this study (hereafter referred to as the team) began to decide which learning outcomes to use for the concept inventories. The goal was to cover all the fundamental statements from the ASMCG.

Early in our efforts, we realized two important issues. First, while immunology is not an explicit part of the ASMCG, it is a common topic in general microbiology courses, and the team determined it needed to be included. We therefore created a fundamental statement and learning outcomes to address this deficit. Second, there are two broad types of microbiology courses, those whose clientele are pre-health professionals (such as pre-nursing students and majors such as Nutrition and Health) and that therefore have a greater emphasis on the medical aspects of microbiology, and those courses that take a more general approach (including majors such as Biology, Microbiology, Biomedical Sciences, etc.), with a greater emphasis on metabolism and diversity of microorganisms. Attempts to serve both constituencies with one concept inventory were difficult, and the optimum solution was to create two concept inventories, with some overlap, to target both audiences. Thus, the working teams were split into the Microbiology Concept Inventory (MCI) and the Microbiology for Health Sciences Concept Inventory (MHSCI). Heather Seitz became director of the MHSCI, while Timothy Paustian remained to direct the MCI, with both teams coordinating their efforts. Table 1 shows the fundamental statements and learning outcomes that were chosen for the MCI.

Identification of commonly held misconceptions

To determine student thinking about the selected concepts and to uncover commonly held misconceptions, we developed a set of true/false (T/F) questions that targeted fundamental statements from the ASMCG. The team of 12

TABLE 1.
MCI fundamental statements mapped to concept inventory questions.

Fundamental Statement	Question
EVOLUTION	
1. Cells, organelles (e.g., mitochondria and chloroplasts), and all major metabolic pathways evolved from early prokaryotic cells.	2
2. Mutations and horizontal gene transfer, with the immense variety of microenvironments, have selected for a huge diversity of microorganisms.	1
3. Human impact on the environment influences the evolution of microorganisms (e.g., emerging diseases and the selection of antibiotic resistance).	7
4. The traditional concept of species is not readily applicable to microbes due to asexual reproduction and the frequent occurrence of horizontal gene transfer.	3
5. The evolutionary relatedness of organisms is best reflected in phylogenetic trees.	2, 23
CELL STRUCTURE AND FUNCTION	
6. The structure and function of microorganisms have been revealed by the use of microscopy (including bright field, phase contrast, fluorescent, and electron).	8
7. Bacteria have unique cell structures that can be targets for antibiotics, immunity, and phage infection.	5, 6, 19
8. Bacteria and Archaea have specialized structures (e.g., flagella, endospores, and pili) that often confer critical capabilities.	4, 6
9. While microscopic eukaryotes (e.g., fungi, protozoa, and algae) carry out some of the same processes as bacteria, many of the cellular properties are fundamentally different.	9
10. The replication cycles of viruses (lytic and lysogenic) differ among viruses and are determined by their unique structures and genomes.	17
METABOLIC PATHWAYS	
11. Bacteria and Archaea exhibit extensive, and often unique, metabolic diversity (e.g., nitrogen fixation, methane production, anoxygenic photosynthesis).	20
12. The interactions of microorganisms among themselves and with their environment are determined by their metabolic abilities (e.g., quorum sensing, oxygen consumption, nitrogen transformations).	13
13. The survival and growth of any microorganism in a given environment depends on its metabolic characteristics.	11
14. The growth of microorganisms can be controlled by physical, chemical, mechanical, or biological means.	5, 10, 12, 19
INFORMATION FLOW AND GENETICS	
15. Genetic variations can impact microbial functions (e.g., in biofilm formation, pathogenicity, and drug resistance).	4
16. Although the central dogma is universal in all cells, the processes of replication, transcription, and translation differ in Bacteria, Archaea, and Eukaryotes.	16
17. The regulation of gene expression is influenced by external and internal molecular cues and/or signals.	15
18. The synthesis of viral genetic material and proteins is dependent on host cells.	17
19. Cell genomes can be manipulated to alter cell function.	15, 16
MICROBIAL SYSTEMS	
20. Microorganisms are ubiquitous and live in diverse and dynamic ecosystems.	7, 13
21. Most bacteria in nature live in biofilm communities.	4, 21
22. Microorganisms and their environment interact with and modify each other.	7, 13
23. Microorganisms, cellular and viral, can interact with both human and non-human hosts in beneficial, neutral, or detrimental ways.	10, 18
IMPACT OF MICROORGANISMS	
24. Microbes are essential for life as we know it and the processes that support life (e.g., in biogeochemical cycles and plant and/or animal microbiota).	7, 18
25. Microorganisms provide essential models that give us fundamental knowledge about life processes.	14
26. Humans utilize and harness microbes and their products.	16
27. Because the true diversity of microbial life is largely unknown, its effects and potential benefits have not been fully explored.	7, 18, 22
IMMUNOLOGY	
The immune system recognizes microbial pathogens and fights against disease.	10

researchers was divided into four subgroups and charged with writing T/F questions based upon selected learning outcomes related to the fundamental statements. The questions were written by individuals in the subgroup and then sequentially reviewed by each subgroup, reviewed by the entire team of 12, and edited when necessary. Question creation was an iterative process, with edits being made both in subcommittee and in the large group to arrive at clear wording that reflected best practices in writing MCQs (34). Each T/F question was accompanied by a free-response prompt for students to explain their reasoning. The resulting T/F test (MCI 1.0) was given at eight colleges and universities (six public and two private), obtaining 743 total responses (Table 2).

Free responses were processed to eliminate unhelpful responses (e.g., I don't know, guess) and then randomized. Enough remaining responses were examined to determine the top three misconceptions that students had on each question. In most cases, between 300 and 360 free responses were scored. Three questions had to be revised due to student responses demonstrating a misunderstanding of the wording of the T/F prompt. For example, in question 12, asking about a change in growth rate after a shift in temperature, the accompanying graph showed the growth rate of a bacterium. The original graph did not have a clear enough difference between temperatures for students to arrive at the correct conclusion. The graph was redrawn to make it more interpretable.

Development of MCQs and development and validation of the concept inventory

Multiple-choice questions were created from each T/F question using the common misconceptions identified from MCI 1.0 to create distractors. In addition, three new T/F questions were created to cover fundamental statements 5 and 21, which were missing in the previous test due to elimination or revision of questions. This set of MCQs became MCI 2.0, which was then administered at colleges and universities listed in Table 2, generating 1,043 usable responses. Simultaneously, an independent four-member review team analyzed MCI 2.0 for accuracy and clarity. The review team also served as experts for the purposes of validation, noting vague wording and identifying the learning objectives that each question assessed independently of the concept inventory design team.

The MCI and MHSCI working teams and the faculty review team came together in a workshop and reviewed the results of MCI 2.0 and the MHSCI. Each question was analyzed for difficulty, item discrimination, and its point-biserial correlation coefficient (r_{pbs}). Most of the question items gave encouraging scores and were kept, with one item, item 7, being removed. In addition, new questions were developed from the additional T/F questions, based upon the misconceptions expressed by the students. Finally, one question (question 14 on MCI 3.0) was taken from the host-pathogen interactions concept inventory (26). Both the MCI and MHSCI were compared to ensure appropriate overlap and differentiation between them. Out of this workshop, MCI 3.0 and the final MHSCI were created.

Administration and evaluation of MCI 3.0

MCI 3.0 was tested with 1,161 students at 10 institutions (Table 2) using a pre- and post-course test approach. Data were combined and the test responses were evaluated using psychometric classic test theory and other methods to determine Ferguson's delta, Cronbach's alpha, item difficulty, item discrimination, and the r_{pbs} (item to total correlation). Analysis of student learning gains was conducted using a normalized learning gain (34). Student performance was also evaluated by subtracting the item difficulty in the

TABLE 2.
Number of responses at colleges where the MCI was tested.

College	College Type	MCI 1.0	MCI 2.0	MCI 3.0
Beloit College	4-year private	10	37	19
Concordia University	4-year private	81	83	10
Iowa State University	4-year public R1	0	0	22
Rogers State University	4-year public	37	0	8
Sam Houston State University	4-year public	64	117	0
University of California – Irvine	4-year public R1	0	0	303
University of Central Oklahoma	4-year public	11	52	22
University of Maryland	4-year public R1	0	156	212
University of North Texas	4-year public R1	35	32	151
University of Wisconsin – Madison	4-year public R1	246	201	143
Virginia Tech	4-year public R1	259	365	271
	Totals	743	1,043	1,161

MCI = microbiology concept inventory. R1 = Research University.

pretest from the item difficulty on the posttest. A negative value indicates the item was found to be less difficult after instruction, which is to be expected.

RESULTS

Design of the MCI

The MCI was developed using the fundamental statements and, when possible, the learning outcomes from the ASMCG for Undergraduate Microbiology (27). Since there are 27 fundamental statements and numerous learning outcomes, many MCI questions were designed to serve multiple fundamental statements. Table 1 maps the MCI questions to their fundamental statements, demonstrating the coverage of the instrument. Some editing and choices had to be made to create a usable-sized test, resulting in less coverage of some fundamental statements.

An effective CI in microbiology will measure a student's understanding of important concepts, as derived from the fundamental statements put forth by the curriculum guidelines, and will provide reliable information about the relative effectiveness of instructional strategies to address those concepts. A good CI will also be both internally consistent and able to justify inferences drawn from the test about student's understanding of each concept measured. These two properties are known as reliability and validity. We expected the MCI to be both reliable and valid.

Test reliability

Reliability is the extent to which a test is repeatable, yielding consistent scores with students of comparable mastery (35). There are a variety of statistical measures that can be used to assess reliability (Table 3). Cronbach's alpha was chosen since it has been used frequently to assess other concept inventories and because it can be used for determining the internal reliability of a test and does not require a retest, which is often difficult to carry out. In such cases, Cronbach's alpha is used to assess scale reliability in lieu of gauging pure test-retest reliability. Cronbach's alpha assesses the reliability or internal consistency of items within a test. The coefficient can range from 0 to 1, where values closer to 1 demonstrate the items are measuring similar underlying concepts, in this case mastery of important concepts in microbiology. It is desirable to have a Cronbach's alpha in the range of 0.65 to 0.8. Measures of the posttest using Cronbach's alpha resulted in a value of $\alpha = 0.705$.

Ferguson's delta measures the discriminatory power of the instrument by measuring the between-person differences of the student scores. Ideally, one finds a broad distribution of test scores, and this is thought to indicate better discrimination. Ferguson's delta is more a measure of the population of students than the test itself, but if a test and population have a Ferguson's delta above 0.90 the test is considered to provide good discriminatory power for the population. Ferguson's delta was 0.96 for the pre-instruction MCI and 0.97 for the post-instruction MCI.

Test validity and discriminatory power

The validity of a test indicates how well each item measures what it is supposed to measure, in this case, how well each question assesses its underlying concept. Item difficulty measures how many students answer an item correctly; it is a simple ratio of the number of correct answers to the number of students. Difficulty can range from 0, no correct answers, to 1, all correct answers. Item difficulty ranged from 0.24 to 0.84 on the posttest (Table 4), with an average difficulty across all questions of 0.54.

Item discrimination (D) measures the degree to which success on an item is indicative of success on the assessment. D is computed by comparing equal-sized high and low scoring groups. The scores of an item of the low-scoring group are subtracted from the high-scoring group and then divided by the size of the group. The range of the index is

TABLE 3.
Statistical measures used to evaluate the MCI.

Name of Test	Function	Recommended Values
Ferguson's delta	Shows how broadly the total scores are distributed over the possible range and measures the discriminatory power of the entire test	0.90 and above is the gold standard; higher is better (38)
Cronbach's alpha	Internal consistency/reliability measure of how well a test addresses different constructs and delivers reliable scores	≥ 0.7 is desirable (36)
Item difficulty	The percentage of students getting items correct	There should be a range of difficulties. Best measured in the posttest (36, 37)
Item discrimination (D_{27})	Compares upper percentile to lower percentile to check how well questions discriminate between strong and weak students	Values should not be negative. Good values are ≥ 0.3 (36)
Point-biserial correlation coefficient (r_{pbs})	Correlates the individual student's performance on a binary test item to their overall performance on the entire test.	Negative values could indicate a defective item, and low values meeting the 0.20 threshold or higher can indicate a question is probing specific knowledge (36, 39)

-1 to 1. If the standard percentile of 27% is used, values of 0.4 and above are considered high and values less than 0.2 are considered low (36). D ranged from 0.25 to 0.61 for the posttest MCI 3.0 (Table 4).

The point-biserial correlation coefficient (r_{pbs}) determines the Pearson correlation between a particular binary item and the whole-test score. It determines the strength of association between two variables, in this case student performance on a question (either correct or incorrect) compared with their score on the entire test. The r_{pbs} ranges from -1 to 1, with a negative value sometimes indicating a defective item, or one that is too general in nature. A value closer to 1 indicates that a particular item is highly associated with the overall test score itself and, if too high, that it may not be probing for specific knowledge being tested on an individual item. A typical range for items with acceptable coefficients is 0.3 to 0.7, and all items should be above 0.2. The r_{pbs} for each item ranged from 0.24 to 0.55 (Table 4). None of the questions were found to be defective and all passed the 0.2 threshold. Often, a low r_{pbs} passing the 0.2 threshold is desirable and shows that a particular question is addressing an underlying concept specifically, in lieu of measuring general test knowledge or test-taking ability.

TABLE 4.
Statistical item tests of the MCI.

Question #	Item Difficulty	Item Discrimination (D)	Point-Biserial Correlation Coefficient (r_{pbs})
1	0.49	0.6	0.48
2	0.74	0.41	0.36
3	0.61	0.4	0.32
4	0.46	0.33	0.28
5	0.72	0.43	0.39
6	0.61	**0.61**	0.49
7	0.42	0.41	0.35
8	0.52	0.39	0.3
9	0.36	0.32	0.28
10	0.54	0.46	0.4
11	**0.24**	0.29	0.3
12	0.58	0.46	0.39
13	0.53	0.31	0.25
14	0.76	0.46	0.45
15	0.77	0.54	**0.55**
16	0.29	0.39	0.33
17	0.27	**0.25**	0.25
18	**0.83**	0.39	0.45
19	0.25	0.44	0.41
20	0.52	0.4	0.33
21	0.73	0.52	0.48
22	0.44	0.28	**0.24**
23	0.7	0.46	0.42
Mean±SD	0.55±0.18	0.42±0.10	0.36±0.09

Bolding indicates high and low values.

This is an excellent way to determine the consistency of individual test items within the overall test. A negative D or r_{pbs} value indicates that weaker students answered a question correctly at a higher rate than stronger students. None of the MCI had negative D or r_{pbs} values, usually observed for defective questions. A comparison of items thought to address the same fundamental concepts (Table 1) showed no strong correlations between any of the items.

Comparison of pretest with posttest and overall learning gains

An important goal of concept inventories is to be able to measure students' mastery of concepts after instruction. While administration of MCI in this research study was focused on the development of the instrument, it was possible to examine the change in understanding of the concepts for the students by comparing the pre- and posttests of MCI 3.0. Comparison of item difficulty and item discrimination is presented in Figure 1. Learning gains and normalized learning gains were calculated for each question, presented in Figures 2 and 3 respectively. In almost all questions, students showed an increase in understanding after instruction on the post-MCI, as expected.

DISCUSSION

Development of the MCI

The strong foundation of the ASMCG and an extended, iterative process greatly facilitated the development of the MCI. Three rounds of survey testing (MCI 1, 2, and 3) allowed the culling of poorly designed questions and the development of replacements. Multiple rounds of surveys allowed the development of alternative questions. The face-to-face meeting between the MCI, the MHSCI, and the review teams, held after MCI 2.0 was tested, accelerated the process. Such meetings are of great use in any CI development.

Mapping of MCI 2.0 to fundamental statements was done independently by two teams in two different manners. First, after the creation of MCI 2.0, the development team checked the alignment of the intended fundamental statements to the questions and verified coverage of the curriculum guidelines. Second, the review team also reverse-engineered MCI 2.0, mapping the questions back to the fundamental statements, as a cross-check to ensure that the questions were clear in addressing their intended statements. MCI 3.0 covers all 27 fundamental statements from the ASMCG.

Distractors for each question were found to be effective as judged by the percentage of students who chose them (Supplementary Table S1). For those students answering a question incorrectly, all distractors were tempting to at least some of the students. In most cases, distractors were chosen by at least 15% of the students who choose an incorrect answer.

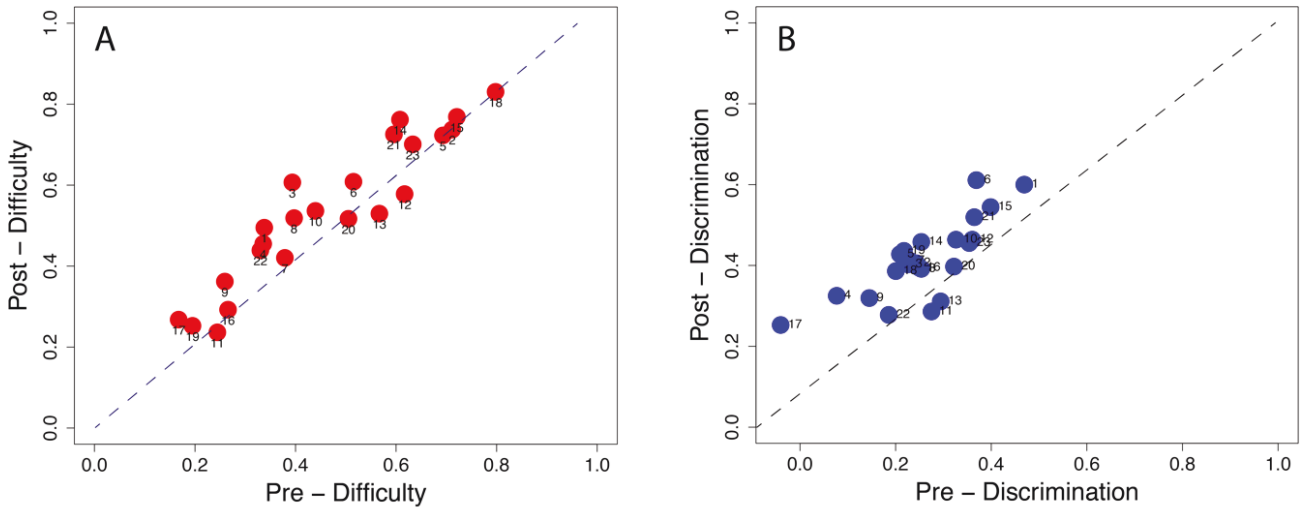

FIGURE 1. Item difficulty and Item discrimination pre- vs. posttest. A total of 1,161 student surveys were used to determine question difficulty and discrimination. The dashed line indicates where each question would land if there was no change in difficulty or discrimination. Measured difficulty of each question decreased after instruction (the difficulty score increased). The discriminatory power of each question increased in the posttest.

Validity and reliability of MCI 3.0

Cronbach's alpha value of 0.705 indicates that the MCI is a reliable concept inventory for assessing understanding of core microbiology concepts. Because Cronbach's alpha assesses the reliability or internal consistency of items within a test, we can determine that the MCI is a reliable instrument that is both multidimensional and internally consistent. Cronbach's alpha of 0.705 demonstrates reliability of the test without exceeding the 0.90 threshold. A value above 0.90 may indicate the items are testing a limited number of concepts, and not a broad understanding of microbiology.

Ferguson's delta value was 0.97 for the posttest, suggesting the MCI does an excellent job of discriminating between students who understand the underlying concepts and those who do not. While this metric is dependent upon the student population, both the number of students tested and the broad range of academic institutions suggest the MCI is applicable to general microbiology courses. Further investigation may benefit from an examination of different student populations, for example, those in two-year versus four-year colleges, those who have taken the course for different reasons, and so forth. Overall, a Ferguson's delta value of 0.97, given the diverse population of students taking the test, not only demonstrates that the test is discriminatory, but that results are generalizable to various populations.

The item difficulty, D, and r_{pbs} for all items in the posttest MCI fell within desirable boundaries. Item response theory suggests that the acceptable range of D is about 0.2 to 0.8 with an average of 0.5 across all items (36). The MCI had a range of 0.24 to 0.83, with an average of 0.55, suggesting a near optimal distribution of difficulties. Question 18 had a high D score (0.83), indicating most students

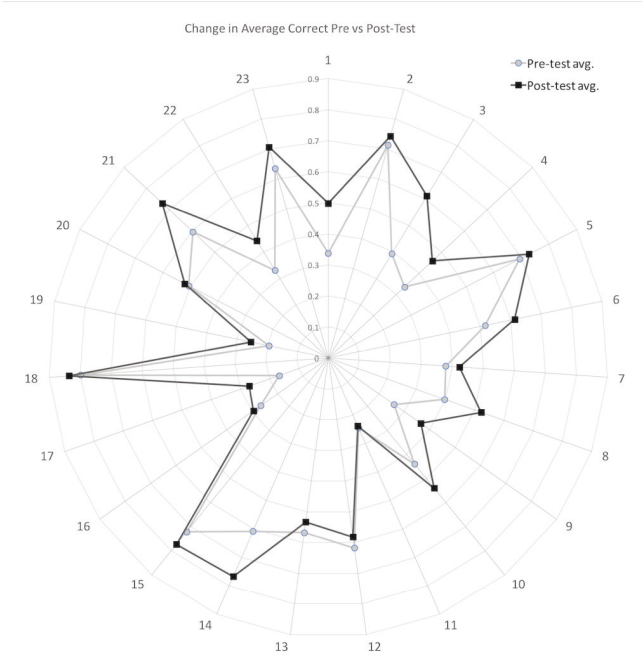

FIGURE 2. Performance by question, pre- vs. posttest. Comparison of the average number of students answering correctly in the Pre-Test (●) vs. the Post-Test (■). Students showed improvement in all but questions 11, 12, and 13.

answered correctly. Item 18 tested student understanding of the roles of the microbiome, and the high D score may be due to the large amount of attention the microbiome has received in the press and the interest the public shows in the subject. Questions 14 and 15 also had higher scores, both dealing with microorganisms' response to the environment. This may be due to students' familiarity with regulation, or

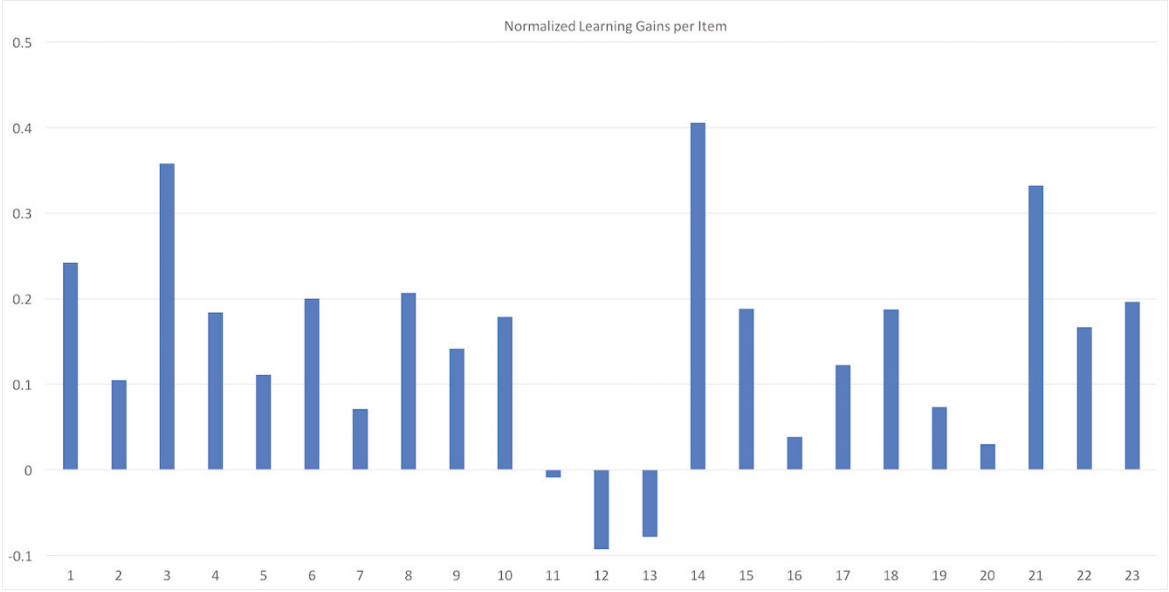

FIGURE 3. Normalized learning gains pre- vs. posttest. The normalized learning gains for each student by question. A total of 1,161 pre- and post- surveys from 10 colleges were analyzed per question. Positive learning gains were found for all but questions 11, 12, and 13.

their ability to guess a sensible answer from the wording of the question. These two questions may be candidates for rewriting to increase difficulty.

D on the posttest ranged from 0.25 to 0.61. All values were above 0.24, indicating the items had good discrimination power between struggling and strong students (37). It is desirable for most items to have D > 0.3, and 20 of 23 items on the MCI meet this standard. Another measure of item discrimination, r_{pbs}, was also used to measure item validity. The r_{pbs} values ranged from 0.24 to 0.55. It is desirable for $r_{pbs} > 0.2$, and all questions met this standard (37), further supporting the validity of the instrument. While some questions were designed to test the same fundamental statements (Table 1), analysis testing the phi coefficient of correlation between these items showed no significant correlations. This is unsurprising, since the fundamental statements are quite broad.

The dashed line in Figure 1A indicates where a test item point would fall if there was no change between the pre- and posttest. A point falling above the line indicates an increase in understanding, while a point falling below the lines indicates a decrease in understanding after instruction. Students taking the pretest had a more difficult time answering the questions, as shown by the lower item difficulty for most of the questions on MCI 3.0. Before instruction, students scored poorly on questions 9, 11, 16, 17, and 19. Question 9 and 16 deal with the differences between prokaryotes and eukaryotes, question 17 is on viral replication, while question 19 focuses on vaccine function. These questions cover topics that are often new concepts to incoming microbiology students, and it is unsurprising that they would initially score poorly. The same effect was observed for discrimination power (Fig. 1B). All but two questions (11 and 13) increased in discriminatory power after instruction (are above the dashed line). This is to be expected since instruction should increase understanding and those students who have mastered the subject matter should differentiate themselves from those who have not. Students taking the pretest all come in with less understanding, and an increase in guessing by all students would decrease the discriminatory power of the MCI.

Measurement of learning gains

A comparison of learning gains showed that for most of the questions, students increased their understanding after instruction. The exceptions were questions 11, 12, and 13. Initial item difficulty scores were low and did not change after instruction. These questions focus on metabolic pathways, an area where students often struggle, and indicate a subject that is a clear target for instructional intervention. The largest learning gains observed were in questions 1, 3, 14, and 21. Questions 1 and 3 are aligned with understanding evolution (Table 1), Question 14 asks about model organisms, and question 21 involves understanding the role of biofilms. These questions show that the largest learning gains seem to be about topics that are not addressed in introductory biology courses or, in the case of evolution, not addressed in the context of microbes. It may be that, across microbiology courses, these gaps in understanding are responsive to instruction. Future work should probe the various types of instruction used in microbiology courses and determine which are most effective.

CONCLUSION

A committee of 15 faculty representing colleges and universities from across the country used the ASMCG to

develop MCI 3.0, consisting of 23 multiple-choice questions, to measure students' broad understanding of microbiology. Whole-test and item psychometric analysis demonstrate that MCI 3.0 is both reliable and valid. Comparison of the pretest with the posttest showed that, after instruction, students generally performed better on MCI 3.0 and the test had high discriminatory power. MCI 3.0 identified learning gains in students and pointed out areas where students still struggle. Those wishing to obtain a copy of MCI 3.0 for their work please go to https://goo.gl/6RTTDS.

SUPPLEMENTAL MATERIALS

Appendix 1: Table S1. Percent of distractors chosen for each question

ACKNOWLEDGMENTS

This research was supported in part by a grant from the National Science Foundation, Division of Undergraduate Education, Conference Proposal #1625772. This material is based upon work supported by the National Institute of Food and Agriculture, USDA, Hatch project IOW04008, Accession # 1010715. The authors would like to acknowledge the support and sage guidance of Amy Chang, director of ASM education. We would also like to thank David L. Popham and Richard W. Seyler, Jr., at Virginia Tech, Hari Kotturi, Clarke Ovrebo, and Carrie Bentley, at University of Central Oklahoma, and Melissa Christopherson, at University of Wisconsin-Madison, for administering the MCI in their General Microbiology courses. Finally, we would like to thank Theodore Muth of Brooklyn College and Christopher Parker of Texas Wesleyan University for their participation on the review team. The authors declare that there are no conflicts of interest.

REFERENCES

1. President's Council of Advisors on Science and Technology (PCAST). 2012. Engage to excel: producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Executive Office of the President, Washington, DC.
2. Handelsman J, Ebert-May D, Beichner R, Bruns P, Chang A, Dehaan R, Gentile J, Lauffer S, Stewart J, Tilghman SM, Wood WB. 2004. Scientific teaching. Source Sci New Ser 304:521–522.
3. Dirks C, Wenderoth MP, Withers M. 2014. Assessment in the college science classroom. W. H. Freeman, United States.
4. Brown PC, Roediger III HL, McDaniel MA. 2014. Make it stick: the science of successful learning. Belknap Press, United States.
5. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. 2014. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci USA 111:8410–8415.
6. Hestenes D, Wells M, Swackhamer G. 1992. Force concept inventory. Phys Teach 30:141.
7. Savinainen A, Scott P. 2002. The force concept inventory: a tool for monitoring student learning. Phys Educ 37:45–52.
8. Sadler PM, Coyle H, Miller JL, Cook-Smith N, Dussault M, Gould RR. 2009. The astronomy and space science concept inventory: development and validation of assessment instruments aligned with the K–12 national science standards. Astron Educ Rev 8:10111.
9. Aslanides JS, Savage CM. 2013. Relativity concept inventory: development, analysis, and results. Phys Rev Spec Top Phys Educ Res 9.
10. Herman GL, Loui M, Zilles C. 2010. Creating the digital logic concept inventory, p 102–106. *In* Proceedings of the 41st ACM Technical Symposium on Computer Science Education.
11. Stone A, Allen K, Rhoads TR, Murphy TJ, Shehab RL, Saha C. 2003. The statistics concept inventory: a pilot study, p T3D1–T3D6. *In* Proceedings – Frontiers in Education Conference, FIE.
12. Epstein J. 2007. Development and validation of the calculus concept inventory, p 165–170. *In* Proceedings of the Ninth International Conference on Mathematics Education in a Global Community.
13. Almstrum VL, Henderson PB, Harvey V, Heeren C, Marion W, Riedesel C, Soh L-K, Tew AE. 2006. Concept inventories in computer science for the topic discrete mathematics. ACM SIGCSE Bull 38:132.
14. Krause S, Birk J, Bauer R, Jenkins B, Pavelich MJ. 2004. Development, testing, and application of a chemistry concept inventory, p 103–107. *In* 34th Annual Frontiers in Education, 2004. FIE 2004.
15. Knudson D. 2006. Biomechanics concept inventory. Percept Mot Skills 103:81–82.
16. D'Avanzo C. 2008. Biology concept inventories: overview, status, and next steps. Bioscience 58:1079.
17. Newman DL, Snyder CW, Fisk JN, Wright LK. 2016. Development of the central dogma concept inventory (CDCI) assessment tool. CBE Life Sci Educ 15(2):ar9.
18. Wang JR. 2004. Development and validation of a two-tier instrument to examine understanding of internal transport in plants and the human circulatory system. Int J Sci Math Educ 2:131–157.
19. Deane T, Nomme K, Jeffery E, Pollock C, Birol G. 2014. Development of the biological experimental design concept inventory (BEDCI). CBE Life Sci Educ 13:540–551.
20. Lin SW. 2004. Development and application of a two-tier diagnostic test for high school students' understanding of flowering plant growth and development. Int J Sci Math Educ 2:175–199.
21. Odom AL, Barrow LH. 1995. Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction. J Res Sci Teach 32:45–61.
22. Smith MK, Wood WB, Knight JK. 2008. The genetics concept assessment: a new concept inventory for gauging student understanding of genetics. CBE Life Sci Educ 7:422–430.

23. Kalas P, O'Neill A, Pollock C, Birol G. 2013. Development of a meiosis concept inventory. CBE Life Sci Educ 12:655–664.
24. Couch BA, Wood WB, Knight JK. 2015. The molecular biology capstone assessment: a concept assessment for upper-division molecular biology students. CBE Life Sci Educ 14:ar10.
25. Stefanski KM, Gardner GE, Seipelt-Thiemann RL. 2016. Development of a lac operon concept inventory (LOCI). CBE Life Sci Educ 15:ar24.
26. Marbach-Ad G, Briken V, El-Sayed NM, Frauwirth K, Fredericksen B, Hutcheson S, Gao LY, Joseph SW, Lee V, McIver KS, Mosser D, Booth Quimby B, Shields P, Song W, Stein DC, Yuan RT, Smith AC. 2009. Assessing student understanding of host-pathogen interactions using a concept inventory. J Microbiol Biol Educ 10:43–50.
27. Horak REA, Merkel S, Chang A. 2015. The ASM curriculum guidelines for undergraduate microbiology: a case study of the advocacy role of societies in reform efforts. J Microbiol Biol Educ 16:100–104.
28. Stevens A, Liao M-K, Merkel S, Chang A. 2014. ASM task force for learning outcomes – ASM general microbiology learning outcome examples. https://www.asm.org/images/Education/FINAL_Learning_Outcomes_w_title_page.pdf
29. Seitz HM, Horak REA, Howard MW, Kluckhohn Jones LW, Muth T, Parker C, Rediske AP, Whitehurst MM. 2017. Development and validation of the microbiology for health sciences concept inventory. J Microbiol Biol Educ 18(3):1–10.
30. Briggs AG, Hughes LE, Brennan RE, Buchner J, Horak REA, Katz-Amburn DS, McDonald AH, Primm TP, Smith AC, Stevens AM, Yung SB, Paustian TD. 2017. Concept inventory development reveals common student misconceptions about microbiology. J Microbiol Biol Educ 18(3):1-9.
31. Adams WK, Wieman CE. 2011. Development and validation of instruments to measure learning of expert-like thinking. Int J Sci Educ 33:1289–1312.
32. Merkel S and the ASM Task Force on Curriculum Guidelines for Undergraduate Biology. 2012. The development of curricular guidelines for introductory microbiology that focus on understanding. J Microbiol Biol Educ 13:32–38.
33. Meltzer DE. 2002. The relationship between mathematics preparation and conceptual learning gains in physics: a possible "hidden variable" in diagnostic pretest scores. Am J Phys 70:1259.
34. Crocker LM, Algina J. 1986. Introduction to classical and modern test theory. Holt, Rinehart, and Winston, New York.
35. Ebel RL. 1954. Procedures for the analysis of classroom tests. Educ Psychol Meas 14:352–364.
36. Bardar EM, Prather EE, Brecher K, Slater TF. 2006. Development and validation of the light and spectroscopy concept inventory. Astron Educ Rev 5:103–113.
37. Ding L, Chabay R, Sherwood B, Beichner R. 2006. Evaluating an electricity and magnetism assessment tool: brief electricity and magnetism assessment. Phys Rev Spec Top Phys Educ Res 2.
38. Ferguson GA. 1949. On the theory of test discrimination. Psychometrika 14:61–68.
39. Anderson DL, Fisher KM, Norman GJ. 2002. Development and evaluation of the conceptual inventory of natural selection. J Res Sci Teach 39:952–978.